

## RESEARCH

## Open Access

# Adaptive matching pursuit with constrained total least squares

Tianyao Huang, Yimin Liu\*, Huadong Meng and Xiqin Wang

**Abstract**

Compressive sensing (CS) can effectively recover a signal when it is sparse in some discrete atoms. However, in some applications, signals are sparse in a continuous parameter space, e.g., frequency space, rather than discrete atoms. Usually, we divide the continuous parameter into finite discrete grid points and build a dictionary from these grid points. However, the actual targets may not exactly lie on the grid points no matter how densely the parameter is grided, which introduces mismatch between the predefined dictionary and the actual one. In this article, a novel method, namely adaptive matching pursuit with constrained total least squares (AMP-CTLS), is proposed to find actual atoms even if they are not included in the initial dictionary. In AMP-CTLS, the grid and the dictionary are adaptively updated to better agree with measurements. The convergence of the algorithm is discussed, and numerical experiments demonstrate the advantages of AMP-CTLS.

**1 Introduction**

A new class of techniques called compressed sampling or compressive sensing (CS) has been widely used recently, due to the fact that CS techniques have shown good performance in different areas such as signal processing, communication and statistics; see, e.g., [1]. Generally, CS finds the sparsest vector  $\mathbf{x}$  from measurements  $\mathbf{y} = \Phi\mathbf{x}$ , where  $\Phi$  is often referred to as *dictionary* with more columns than rows, and each column of the dictionary is called an *atom* or a *basis*.

Matching pursuit (MP) is a set of popular greedy approaches to compressive sensing. The basic idea is to sequentially find the support set of  $\mathbf{x}$  and then project on the selected atoms. The atoms selected in the support set are mainly determined by correlations between atoms and the regularized measurements [2]. MP methods include standard MP [3], and several other examples, such as orthogonal matching pursuit (OMP) [4], regularized OMP (ROMP) [5], stage-wise OMP (StOMP) [6], compressive sampling matching pursuit (CoSaMP) [7] and subspace pursuit (SP) [2].

These MP methods [2-7] do not consider the off-grid problem in grid-based CS approaches. In some applications of CS, such as harmonic retrieval and radar signal processing (e.g., range profiling [8,9], direction of arrival

estimate [10-12]), we usually divide a continuous parameter space into discrete grid points to generate the dictionary. For example, in harmonic retrieval, frequency space is divided and dictionary is a discrete Fourier transform (DFT) matrix. The off-grid problem emerges when the actual frequencies are placed off the predefined grid. The mismatch between the predefined and actual atom can lead to performance degradation in sparse recovery (e.g., [13-15]).

The grid misalignment problem in CS has recently received growing interest. The sensitivity of CS to the mismatch between the predefined and actual atoms is studied in [13]; however, the focus of that article is mainly on mismatch analysis rather than development of an algorithm. Cabrera et al. [16] and Zhu et al. [14], respectively, provided an iterative re-weighted (IRW)-based and a Lasso-based method to recover an unknown vector considering the atom misalignment, whereas we focus on MP methods in this article. Compared with IRW or Lasso, MP methods greedily find the support set and greatly reduce the dimension of the CS problem; thus, they have an advantage in computability. Gabriel [17] proposed best basis compressive sensing in a tree-structured dictionary, but some dictionaries (e.g., DFT matrix) do not possess a tree structure.

To alleviate the off-grid problem in matching pursuit, we developed adaptive matching pursuit with constrained total least squares (AMP-CTLS). In AMP-CTLS, we

\* Correspondence: [yiminliu@tsinghua.edu.cn](mailto:yiminliu@tsinghua.edu.cn)  
Department of Electronic Engineering, Tsinghua University, Beijing, China

model the grid as an unknown parameter, and adaptively search for the best one. We choose harmonic retrieval to demonstrate the performance of AMP-CTLS. The algorithm can also be applied to jointly estimate range and velocity in randomized step frequency (RSF) radar. Note that in the RSF scenario range-velocity estimation is hard to be directly solved by subspace-based methods, e.g., Capon's method, MUSIC and ESPRIT [18]. Since only one snapshot data is available in RSF radar, to obtain the covariance matrix these subspace-based methods need to apply smoothing method, which requires uniform and linear condition [18]. However, this condition is not satisfied in the case of random frequency model.

This article is structured as follows. Section 2 introduces grid-based CS and outlines the procedures of AMP-CTLS. In Sections 3 and 4, we discuss the implementation of AMP-CTLS in harmonic retrieval and RSF radar, respectively. In Section 5, numerical examples are presented to illustrate merits of AMP-CTLS. Section 6 is dedicated to a brief conclusion. *Notations:*  $(\cdot)^H$  denotes conjugate transpose matrix;  $(\cdot)^T$  transpose matrix;  $(\cdot)^*$  conjugate matrix;  $(\cdot)^\dagger$  pseudo-inverse matrix;  $\mathbf{I}_L/\mathbf{0}_L$  the  $L \times L$  identity/zero matrix;  $\|\cdot\|_2$  the  $\ell_2$  norm;  $\{\cdot\}$  denotes a set;  $|\cdot|$  the absolute value of a complex number or the cardinality of a set;  $(\cdot)_\Lambda$  denotes elements/columns indexed in the set  $\Lambda$  of a vector/matrix;  $\text{supp}(\cdot)$  is the support set of a vector, that is, the indices of the nonzero elements in the vector;  $\text{Re}(\cdot)$  the real part of a complex number;  $\otimes$  denotes the right Kronecker product [19]; and  $E[\cdot]$  denotes the expectation of a random variable.

## 2 Grid-based CS and the AMP-CTLS algorithm

The signal model of grid-based CS is introduced in Section 2.1. We combine the greedy idea of MP methods and the constrained total least squares (CTLS) technique [20], and thus produce AMP-CTLS to alleviate the off-grid problem. In AMP-CTLS, the grid is cast as an unknown parameter, and is jointly estimated together with  $\mathbf{x}$ . In Section 2.2, the framework of AMP-CTLS is given. Section 2.3 is dedicated to the iterative joint estimator (IJE) algorithm, which is implemented in AMP-CTLS. In the IJE algorithm, the CTLS technique is used, which is presented in Section 2.4. Section 2.5 summarizes the entire procedure of AMP-CTLS. In Section 2.6, the convergence of IJE is analyzed.

### 2.1 Grid-based CS

CS promises efficient recovery of sparse signals. In many applications, signals are sparse in a continuous parameter space rather than finite discrete atoms. Usually, we divide the continuous parameter into discrete grid points and cast the problem as a grid-based CS model:

$$\mathbf{y} = \Phi(\mathbf{g})\mathbf{x} + \mathbf{w}, \quad (1)$$

where  $\mathbf{y} \in \mathbb{C}^{M \times 1}$  and  $\mathbf{w} \in \mathbb{C}^{M \times 1}$  are measurement vector and white Gaussian noise (WGN) vector, respectively.  $\mathbf{x} \in \mathbb{C}^{N \times 1}$  is to be learned.  $\mathbf{g} \in \mathbb{C}^{N \times 1}$  are discrete grid points  $\mathbf{g} = [g_1, g_2, \dots, g_N]$ .  $\Phi(\mathbf{g}) \in \mathbb{C}^{M \times N}$  is built from  $\mathbf{g}$ ,  $\Phi(\mathbf{g}) = [\phi(g_1), \phi(g_2), \dots, \phi(g_N)]$ , and the mapping  $\mathbf{g} \rightarrow \Phi$  is known. For example, to recover a frequency sparse signal, we grid the frequency space into discrete frequency points  $\mathbf{g} = [0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}]^T$ .  $\Phi$  is a DFT matrix, of which the  $m$ th-row,  $n$ th-column element is  $\exp(j2\pi \frac{m}{N} n)$ . However, the signal is only sparse in the DFT atoms if all of the sinusoids are exactly at the pre-defined grid points [13]. In some cases, no matter how densely we grid the frequency space, the sinusoids could be off-grid, which saps the performance of CS methods [13].

### 2.2 Main idea of AMP-CTLS

The off-grid problem usually emerges because we do not often have enough priori knowledge to generate a perfect grid to guarantee that all of the signals exactly lie on grid points. Thus, we cast the grid as an unknown parameter, and search for the best grid  $\mathbf{g}$  as well as the sparsest  $\mathbf{x}$  by solving the optimum problem:

$$\hat{\mathbf{x}}, \hat{\mathbf{g}} = \arg \min_{\mathbf{x}, \mathbf{g}} \|\mathbf{x}\|_0, \text{ s.t. } \|\mathbf{y} - \Phi(\mathbf{g})\mathbf{x}\|_2^2 \leq \eta, \quad (2)$$

where  $\eta$  is the noise power. Equation (2) is similar to that used in traditional MP methods [2-7], except that we recover  $\mathbf{x}$  and simultaneously estimate the grid. In most cases, solving (2) is a complex non-linear optimum problem. In this article, an iterative method is introduced.

AMP-CTLS inherits the greedy idea from MP methods, which use correlations to iteratively find the support set. In each iteration, one or more atoms are added into the support set. Suppose the support set is obtained as  $\Lambda^{(k)}$  after the  $k$ th iteration, and denote the corresponding grid points as  $\hat{\mathbf{g}}_\Lambda^{(k)}$ . In traditional MP methods [2-7],  $\mathbf{x}_\Lambda$  is estimated by solving a least squares problem. In AMP-CTLS, considering the off-grid problem, we jointly search for  $\mathbf{x}_\Lambda$  and the best grid points in the neighboring continuous region of  $\hat{\mathbf{g}}_\Lambda^{(k)}$  via (3), in which we minimize norm of the *residual error*, which is defined as  $\mathbf{r} = \mathbf{y} - \Phi(\mathbf{g}_\Lambda)\mathbf{x}_\Lambda$ .

$$\hat{\mathbf{x}}_\Lambda^{(k+1)}, \hat{\mathbf{g}}_\Lambda^{(k+1)} = \arg \min_{\mathbf{x}_\Lambda, \mathbf{g}_\Lambda} \|\mathbf{y} - \Phi(\mathbf{g}_\Lambda)\mathbf{x}_\Lambda\|_2^2 \quad (3)$$

We develop the iterative joint estimator (IJE) algorithm to solve (3), which is detailed in ensuing section.

### 2.3 IJE algorithm

It is difficult to find an analytical solution to (3). The IJE algorithm is devised to seek a numerical solution. Given initial grid points  $\hat{\mathbf{g}}_{\Lambda}(0)$ , IJE searches for the best grid points  $\mathbf{g}_{\Lambda}$  in the neighborhood of  $\hat{\mathbf{g}}_{\Lambda}(0)$ . The mismatch of the grid is denoted as  $\Delta\mathbf{g}_{\Lambda} = \mathbf{g}_{\Lambda} - \hat{\mathbf{g}}_{\Lambda}(0) = [\Delta g_1, \dots, \Delta g_{|\Lambda|}]^T$ . IJE includes three steps: calculate the estimation of the mismatch,  $\widehat{\Delta\mathbf{g}}_{\Lambda}$ ; update the grid with  $\widehat{\Delta\mathbf{g}}_{\Lambda}$ ; and estimate  $\mathbf{x}_{\Lambda}$  with projection onto the new grid points. These three steps are executed iteratively to pursue more accurate results. To distinguish from iterations in search for the support set in (3), we denote  $l$  as the counter of loops in IJE; thus, IJE is expressed as follows:

$$\widehat{\Delta\mathbf{g}}_{\Lambda}(l), \mathbf{x}_{\text{CTLS}} = \arg \min_{\Delta\mathbf{g}_{\Lambda}, \mathbf{x}_{\Lambda}} C_{\text{CTLS}}, \quad (4)$$

$$\hat{\mathbf{g}}_{\Lambda}(l+1) = \hat{\mathbf{g}}_{\Lambda}(l) + \widehat{\Delta\mathbf{g}}_{\Lambda}(l), \quad (5)$$

$$\hat{\mathbf{x}}_{\Lambda}(l+1) = \arg \min_{\mathbf{x}_{\Lambda}} \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_{\Lambda}(l+1))\mathbf{x}_{\Lambda}\|_2^2. \quad (6)$$

In (4), CTLS technique is applied to simultaneously search for the mismatch  $\Delta\mathbf{g}_{\Lambda}$  and  $\mathbf{x}_{\Lambda}$ , and  $\widehat{\Delta\mathbf{g}}_{\Lambda}(l)$  and  $\mathbf{x}_{\text{CTLS}}$  are the results.  $C_{\text{CTLS}}$  denotes the penalty function of CTLS, which is detailed in Section 2.4. Since (6) is a linear least squares problem, the closed-form solution is

$$\hat{\mathbf{x}}_{\Lambda}(l+1) = (\Phi(\hat{\mathbf{g}}_{\Lambda}(l+1)))^{\dagger} \mathbf{y} = (\Phi^H \Phi)^{-1} \Phi^H \mathbf{y}. \quad (7)$$

The loops are terminated when the norm of residual error is scarcely reduced.

### 2.4 CTLS technique

Traditional MP methods [2-7] apply least squares to calculate amplitudes of  $\mathbf{x}_{\Lambda}$  after finding the support set. When there are off-grid signals, mismatches occur in the dictionary; thus, we replace the least squares model with total least squares (TLS) criterion, which is appropriate to deal with the fitting problem when perturbations exist in both the measurement vector and in the dictionary [21]. Since the dictionary mismatches are constrained by errors of grid points, we introduce the constrained total least squares (CTLS) technique [20] in AMP-CTLS to jointly estimate the grid point errors and  $\mathbf{x}_{\Lambda}$ , i.e., solving (4). It has been proved that CTLS is a constrained space state maximum likelihood estimator [20].

Suppose that we obtain the estimate of grid points as  $\hat{\mathbf{g}}_{\Lambda}(l)$  after  $l$ th IJE iteration. Assume that the mismatch  $\Delta\mathbf{g}_{\Lambda}$  is significantly small; thus we can approximate the

perfect dictionary  $\Phi(\mathbf{g}_{\Lambda})$  as a linear combination of the mismatch  $\Delta\mathbf{g}$  with Taylor expansion:

$$\Phi(\mathbf{g}_{\Lambda}) = \Phi(\hat{\mathbf{g}}_{\Lambda}(l)) + \sum_{i=1}^{|\Lambda|} \mathbf{R}_i(\hat{\mathbf{g}}_{\Lambda}(l)) \Delta g_i + \sum_{i=1}^{|\Lambda|} o(\Delta g_i^2), \quad (8)$$

where  $\mathbf{R}_i \in \mathbb{C}^{M \times |\Lambda|}$  is

$$\mathbf{R}_i(\hat{\mathbf{g}}_{\Lambda}(l)) = \left. \frac{\partial \Phi(\mathbf{g}_{\Lambda})}{\partial g_i} \right|_{\mathbf{g}_{\Lambda}=\hat{\mathbf{g}}_{\Lambda}(l)} \quad (9)$$

and  $o(\cdot)$  denotes higher order terms. For simplicity, in this section we ignore the iteration counter in the notations, and  $\mathbf{R}_i(\hat{\mathbf{g}}_{\Lambda}(l))$ ,  $\Phi(\hat{\mathbf{g}}_{\Lambda}(l))$  are, respectively, simplified as  $\mathbf{R}_i$ ,  $\Phi_{\Lambda}$ . Neglect  $o(\Delta g_i^2)$  and the signal model in (1) is replaced by:

$$\mathbf{y} = \left( \Phi_{\Lambda} + \sum_{i=1}^{|\Lambda|} \mathbf{R}_i \Delta g_i \right) \mathbf{x}_{\Lambda} + \mathbf{w}. \quad (10)$$

CTLS models  $\Delta\mathbf{g}_{\Lambda}$  as an unknown random perturbation vector. The grid misalignment and the noise vector are combined into a  $(M + |\Lambda|)$ -dimensional vector  $\mathbf{v} = [(\Delta\mathbf{g}_{\Lambda})^T, \mathbf{w}^T]^T$ , and CTLS aims at minimizing  $\|\mathbf{v}\|_2^2$ . It has been proved that CTLS is a constrained space state maximum likelihood estimator if  $\mathbf{v}$  is a WGN vector [20]. Thus, we first whiten  $\mathbf{v}$ . Assume that  $\Delta\mathbf{g}_{\Lambda}$  is independent of  $\mathbf{w}$ . The covariance matrix of  $\Delta\mathbf{g}_{\Lambda}$  is  $\mathbf{C}_g = E[\Delta\mathbf{g}_{\Lambda}(\Delta\mathbf{g}_{\Lambda})^H] \in \mathbb{C}^{|\Lambda| \times |\Lambda|}$ .  $\mathbf{D} \in \mathbb{C}^{|\Lambda| \times |\Lambda|}$  obeys  $\mathbf{C}_g^{-1} = \mathbf{D}^H \mathbf{D}$ . The variance of white noise  $\mathbf{w}$  is  $\sigma_w^2$ . We denote an unknown normalized vector  $\mathbf{u} \in \mathbb{C}^{(M+|\Lambda|) \times 1}$  as (11); thus,  $\mathbf{u}$  is a WGN vector.

$$\mathbf{u} = \begin{bmatrix} \mathbf{D} \Delta\mathbf{g}_{\Lambda} \\ \frac{1}{\sigma_w} \mathbf{w} \end{bmatrix} \quad (11)$$

Minimize the penalty function  $C_{\text{CTLS}} = \|\mathbf{u}\|_2^2$  and (4) is detailed as follows:

$$\hat{\mathbf{u}}, \mathbf{x}_{\text{CTLS}} = \arg \min_{\mathbf{u}, \mathbf{x}_{\Lambda}} \|\mathbf{u}\|_2^2, \quad (12)$$

$$\text{s.t. } -\mathbf{y} + \left( \Phi_{\Lambda} + \sum_i^{|\Lambda|} \mathbf{R}_i \Delta g_i \right) \mathbf{x}_{\Lambda} + \mathbf{w} = \mathbf{0}. \quad (13)$$

The constraint condition (13) can be rewritten as:

$$\text{s.t. } -\mathbf{y} + \Phi_{\Lambda} \mathbf{x}_{\Lambda} + \mathbf{W}_x \mathbf{u} = \mathbf{0}, \quad (14)$$

where  $\mathbf{W}_x = [\mathbf{H} \ \sigma_w \mathbf{I}_M] \in \mathbb{C}^{M \times (|\Lambda|+M)}$ .  $\mathbf{H} \in \mathbb{C}^{M \times |\Lambda|}$  is defined as

$$\mathbf{H} = \mathbf{G}(\mathbf{D}^{-1} \otimes \mathbf{I}_{|\Lambda|})(\mathbf{I}_{|\Lambda|} \otimes \mathbf{x}_\Lambda), \quad (15)$$

where  $\mathbf{G} = [\mathbf{R}_1, \dots, \mathbf{R}_{|\Lambda|}] \in \mathbb{C}^{M \times |\Lambda|^2}$ . The equivalence between (13) and (14) is proved as follows:

$$\begin{aligned} \left( \sum_{i=1}^{|\Lambda|} \mathbf{R}_i \Delta \mathbf{g}_i \right) \mathbf{x}_\Lambda &= \mathbf{G}(\Delta \mathbf{g}_\Lambda \otimes \mathbf{I}_{|\Lambda|}) \mathbf{x}_\Lambda \\ &= \mathbf{G}(\mathbf{D}^{-1} \otimes \mathbf{I}_{|\Lambda|})(\mathbf{D} \Delta \mathbf{g}_\Lambda \otimes \mathbf{I}_{|\Lambda|}) \mathbf{x}_\Lambda \quad (16) \\ &= \mathbf{G}(\mathbf{D}^{-1} \otimes \mathbf{I}_{|\Lambda|})(\mathbf{I}_{|\Lambda|} \otimes \mathbf{x}_\Lambda) \mathbf{D} \Delta \mathbf{g}_\Lambda \\ &= \mathbf{H} \mathbf{D} \Delta \mathbf{g}_\Lambda. \end{aligned}$$

When  $\mathbf{W}_x$  is of full-row rank, the optimum problem (12, 14) are equivalent to (17)-(19), which has been proved in [20].

$$\mathbf{x}_{\text{CTLS}} = \min_{\mathbf{x}_\Lambda} \|\mathbf{W}_x^\dagger (\mathbf{y} - \Phi_\Lambda \mathbf{x}_\Lambda)\|_2^2 \quad (17)$$

$$\hat{\mathbf{u}} = \mathbf{W}_x^\dagger (\mathbf{y} - \Phi_\Lambda \mathbf{x}_\Lambda)|_{\mathbf{x}=\mathbf{x}_{\text{CTLS}}} \quad (18)$$

$$\mathbf{W}_x^\dagger = \mathbf{W}_x^H (\mathbf{W}_x \mathbf{W}_x^H)^{-1} \quad (19)$$

It is quite difficult to obtain analytical solution to (17). A complex version of Newton method is developed in [20], which is presented in Appendix 1. Initial value of  $\mathbf{x}_\Lambda$  required in Newton's method for (17) can be given as:

$$\mathbf{x}_{\text{ini}} = \Phi_\Lambda^\dagger \mathbf{y} = (\Phi_\Lambda^H \Phi_\Lambda)^{-1} \Phi_\Lambda^H \mathbf{y}. \quad (20)$$

$\widehat{\Delta \mathbf{g}}_\Lambda$  is extracted from  $\hat{\mathbf{u}}$  via  $\widehat{\Delta \mathbf{g}}_\Lambda = [\mathbf{D}^{-1} \mathbf{0}_N] \hat{\mathbf{u}}$ , thus (4) is solved. The sketch of CTLS is given in Algorithm 1. As the authors' best knowledge, the convergence guarantees for this Newton method are still open question.

## 2.5 Sketch of AMP-CTLS

Similarly to traditional MP methods [2-7], AMP-CTLS first greedily finds the support set. Then AMP-CTLS adaptively optimizes the grid points indexed in the support set. In this article, we imitate the greedy approach of OMP, in which only one atom is added to the support set in each iteration. If the number of atoms is known, terminate the iterations when the cardinality of the support set reaches the pre-specified number. If it is not known, we can apply some other successfully used stopping criterions, e.g., norm of residual being below a threshold [22]. A sketch of AMP-CTLS is presented in Algorithm 2.

## 2.6 Convergence of the IJE algorithm

Here, we analyze convergence of the IJE algorithm. Assume that the mapping  $\mathbf{g}_\Lambda \rightarrow \Phi(\mathbf{g}_\Lambda)$  is linear, which means

$$\Phi(\mathbf{g}_\Lambda + \Delta \mathbf{g}_\Lambda) = \Phi(\mathbf{g}_\Lambda) + \mathbf{G}(\Delta \mathbf{g}_\Lambda \otimes \mathbf{I}_{|\Lambda|}), \quad (21)$$

and  $\mathbf{G}$  should be a constant matrix.

**Proposition.** If the measurement  $\mathbf{y}$  is perturbed by WGN and (21) is obeyed, IJE monotonically reduces values of the penalty function in (3). The estimates of  $\mathbf{x}_\Lambda$  and  $\mathbf{g}_\Lambda$  satisfy:

$$\|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l)) \hat{\mathbf{x}}_\Lambda(l)\|_2^2 \geq \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l+1)) \hat{\mathbf{x}}_\Lambda(l+1)\|_2^2. \quad (22)$$

**Proof.** Define a penalty function as follows:

$$f_p(\Delta \mathbf{g}_\Lambda, \mathbf{x}_\Lambda) = \sigma_w^2 \|\mathbf{u}\|_2^2 = \sigma_w^2 (\Delta \mathbf{g}_\Lambda)^H \mathbf{C}_g^{-1} \Delta \mathbf{g}_\Lambda + \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l)) \mathbf{x}_\Lambda - \mathbf{G}(\Delta \mathbf{g}_\Lambda \otimes \mathbf{I}_{|\Lambda|}) \mathbf{x}_\Lambda\|_2^2; \quad (23)$$

thus,  $\widehat{\Delta \mathbf{g}}_\Lambda(l)$  and  $\mathbf{x}_{\text{CTLS}}$  are obtained by solving

$$\widehat{\Delta \mathbf{g}}_\Lambda(l), \mathbf{x}_{\text{CTLS}} = \arg \min_{\Delta \mathbf{g}_\Lambda, \mathbf{x}_\Lambda} f_p(\Delta \mathbf{g}_\Lambda, \mathbf{x}_\Lambda), \quad (24)$$

which is the same as (4), for  $\sigma_w^2$  is a constant. Thus, it is satisfied that

$$f_p(\widehat{\Delta \mathbf{g}}_\Lambda(l), \mathbf{x}_{\text{CTLS}}) \leq f_p(\mathbf{0}, \hat{\mathbf{x}}_\Lambda(l)) = \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l)) \hat{\mathbf{x}}_\Lambda(l)\|_2^2. \quad (25)$$

Substitute (5), (21) into  $f_p(\widehat{\Delta \mathbf{g}}_\Lambda(l), \mathbf{x}_{\text{CTLS}})$ , and note that  $\mathbf{C}_g^{-1}$  is a positive definite matrix; thus,

$$\begin{aligned} f_p(\widehat{\Delta \mathbf{g}}_\Lambda(l), \mathbf{x}_{\text{CTLS}}) &= \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l+1)) \mathbf{x}_{\text{CTLS}}\|_2^2 + \sigma_w^2 (\widehat{\Delta \mathbf{g}}_\Lambda(l))^H \mathbf{C}_g^{-1} \widehat{\Delta \mathbf{g}}_\Lambda(l) \\ &\geq \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l+1)) \mathbf{x}_{\text{CTLS}}\|_2^2 \\ &\geq \|\mathbf{y} - \Phi(\hat{\mathbf{g}}_\Lambda(l+1)) \hat{\mathbf{x}}_\Lambda(l+1)\|_2^2, \end{aligned} \quad (26)$$

where the last inequality is taken from (6). The inequalities in (25) and (26) are transformed to equalities if and only if  $\widehat{\Delta \mathbf{g}}_\Lambda(l) = \mathbf{0}$ .  $\square$

For simplicity, we assume that the transform  $\Phi(\mathbf{g}_\Lambda)$  is linear. In some practical applications like harmonic retrieval, linearity is not strictly guaranteed. However, when atom mismatch  $\Delta \mathbf{g}$  is significantly small, the higher order errors due to Taylor expansion (8) are ignorable, and (21) is approximately satisfied. Numerical examples are performed in Section 5, which demonstrate the convergence of the proposed algorithm in the case of harmonic retrieval.

## 3 Application in the harmonic retrieval

In this section, we apply AMP-CTLS in harmonic retrieval. In Section 3.1, the signal model of harmonic retrieval is presented and adverse effects of MP approaches [2-7] in harmonic retrieval is discussed. In Section 3.2, we detail the implementation of AMP-CTLS in harmonic retrieval.

### 3.1 Signal model of harmonic retrieval

Consider a complex sinusoidal signal

$$\gamma_m = \sum_{k=1}^K \alpha_k \exp(j2\pi f_k m) + w_m, \quad (27)$$



where  $y_m$  is the  $m$ th measurement, and  $w_m$  is the  $m$ th noise,  $m = 0, 1, \dots, M - 1$ . There are  $K$  sinusoids, and amplitude  $\alpha_k$ , frequency  $f_k$  of the  $k$ th sinusoid are unknown parameters. When the sinusoids are sparse, i.e.,  $K \ll M$ , harmonic retrieval problem can be solved by grid-based CS approaches. Divide the digital frequency  $f \in [0, 1)$  into  $N$  grid points  $\mathbf{g} = [g_1, g_2, \dots, g_N]^T$ . When all frequencies are exactly at grid points, rewrite (27) as

$$y_m = \sum_{n=1}^N x_n \cdot \exp(j2\pi g_n m) + w_m, \quad (28)$$

where  $g_n$  is the frequency of the  $n$ th grid point and

$$x_n = \begin{cases} \alpha_k, & \text{the } k\text{th sinusoid is present at } n\text{th grid point,} \\ 0, & \text{no sinusoid is present at } n\text{th grid point.} \end{cases} \quad (29)$$

Rewrite (28) in matrix form as

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{w}, \quad (30)$$

where the  $m$ th-row,  $n$ th-column element of  $\Phi$  is of the form  $\varphi(m, n) = \exp(j2\pi g_n m)$ . Apply CS methods to seek the sparsest solution of (30). Then, estimates of the frequencies and amplitudes are obtained with the indices and magnitudes of nonzero coefficients in  $\mathbf{x}$ , respectively. The sparsest solution can be obtained with computational MP methods, which greedily minimize the  $\ell_0$  norm. It can also be obtained by minimizing the  $\ell_1$  norm [23], the quasi-norm [24,25] or the  $\ell_p$ ,  $p \leq 1$  p-norm-like diversity [26].

We focus on MP methods in this article for the high computation efficiency. However, conventional MP methods [2-7] suffer from performance degradation if the frequency space is not perfectly grided. When the frequency is sparsely divided, sinusoids may lie off the grid points, and accuracy of frequency estimates is limited by the gap between neighboring grid points. MP methods iteratively search for the sinusoids. If an off-grid sinusoid emerges, the energy of this sinusoid can not be totally canceled and performs as an interference in the next iterations. The leakage of the energy may mask the weak sinusoids. On the other hand, if the frequency space is densely divided, correlations between atoms are enhanced [27], which also reduces the performance of MP methods. Especially in those MP methods that select multiple atoms into the support set in a single iteration, e.g., CoSaMP, SP, ROMP, and StOMP, highly correlated atoms could be chosen in a same iteration, which impairs the numerical stability of projection onto the adopted atoms.

### 3.2 Harmonic retrieval with AMP-CTLS

The AMP-CTLS algorithm can be applied for harmonic retrieval. AMP-CTLS adaptively finds the atoms and

recovers the sinusoids. In those MP approaches with constant predefined atoms, frequency estimates are discrete values, depending on grid points. In AMP-CTLS frequency estimates are continuous, since estimates of the grid misalignments are continuous. In this section, we adjust two steps of AMP-CTLS presented in Section 2 to better fit the harmonic retrieval problem.

Calculate the  $\mathbf{R}$  matrix in (9). According to (9), the  $m$ th-row,  $i$ th-column element of the  $\mathbf{R}_i$  is expressed as follows:

$$\mathbf{R}_i(m, i) = \exp(j2\pi g_i m) \cdot j2\pi m. \quad (31)$$

Elements in other columns are all zeros.

Adjust the grid-updating formula in (5). In CTLS as presented in Section 2.4, the grid misalignment  $\Delta \mathbf{g}_\Lambda$  is assumed to be a complex vector; therefore, the estimate  $\widehat{\Delta \mathbf{g}_\Lambda}$  is complex. However, frequency grid points are restrained to be real, so regularization  $\Delta \mathbf{g}_\Lambda = (\Delta \mathbf{g}_\Lambda)^*$  should be added to (12) in the case of harmonic retrieval. Unfortunately, the solver becomes complex, which is derived in Appendix 2. For simplicity, (5) is replaced with (32) to approximatively update the grid points:

$$\hat{\mathbf{g}}_\Lambda(l+1) = \hat{\mathbf{g}}_\Lambda(l) + \text{Re}(\widehat{\Delta \mathbf{g}_\Lambda}(l)). \quad (32)$$

## 4 Application in RSF radar

AMP-CTLS can also be applied in randomized step frequency (RSF) radar. RSF radar can improve the range-velocity resolution and avoid range-velocity coupling problems [28,29]. However, RSF radar suffers from the sidelobe pedestal problem, which results in small targets being masked by noise-like components due to dominant targets [29]. Our problem of interest is to recover small targets. When the observed scene is sparse, i.e., only few targets exists, we can use sparse recovery to exploit the sparseness [9]. AMP-CTLS relieves the sidelobe pedestal problem in RSF radar and recovers small targets well.

Correlation-matrix-based spectral analysis methods, e.g., MUSIC, ESPRIT [18], are hard to be directly utilized in range-Doppler estimation in RSF scheme. Since only one snapshot of radar data is available and radar echoes from different scatterers are coherent, smoothing technique is invoked to obtain a full rank correlation matrix [30]. Smoothing method requires that the array is uniform and linear [18]. However, in RSF radar, the echoes are determined by a random permutation of integers, see (34); thus, the uniform and linear condition is not satisfied, which restricts application of correlation-matrix-based methods.

We discuss a specific example of RSF radar, in which the waveform is a monotone pulse signal and the

frequency of the  $m$ th pulse is  $f_0 + C_m \delta f$ ,  $m = 0, 1, \dots, M - 1$ , where  $f_0$  is carrier frequency and  $\delta f$  is frequency step size.  $C_m$  is a random permutation of integers from 0 to  $M - 1$ . The  $m$ th echo of radar can be expressed as (see [8,28,29]):

$$y_m = \sum_{k=1}^K \alpha_k s_m(p_k, q_k) + w_m, \quad (33)$$

$$s_m(p, q) = \exp(-j2\pi C_m p - j2\pi m(1 + C_m \delta f/f_0)q), \quad (34)$$

where  $w_m$  is noise in the  $m$ th echo.  $K$  denotes the number of targets and  $k$  denotes  $k$ th target.  $\alpha_k$ ,  $p_k$ , and  $q_k$  are to be learned.  $\alpha_k$  presents the scattering intensity.  $p_k \in [0, 1)$  and  $q_k \in [0, 1)$  are determined by range and radial velocity of the  $k$ th target, respectively. Note that in (34) the echo is simultaneously related to the sequence  $m$  and the random integer  $C_m$ .

Divide  $p$  space into  $C$  grid points  $p_c = c/C$ ,  $c = 0, 1, \dots, C - 1$ . Divide  $q$  space into  $D$  grid points  $p_d = d/D$ ,  $d = 0, 1, \dots, D - 1$ . Rewrite (33) as:

$$y = \Phi(p, q)x + w, \quad (35)$$

where the  $m$ th-row,  $(c + dC)$ th-column element of  $\Phi$  ( $p, q$ )  $\in \mathbb{C}^{M \times CD}$  is  $s_m(p_c, q_d)$ .

AMP-CTLS is implemented to solve (35). First, we find the support set  $\Lambda$  and then use IJE and CTLS to adjust the grid points, though CTLS described in Section 2.4 requires modification. The grid misalignment vector consists of two parts:  $p$  mismatch  $\Delta p_\Lambda \in \mathbb{R}^{|\Lambda| \times 1}$  and  $q$  mismatch  $\Delta q_\Lambda \in \mathbb{R}^{|\Lambda| \times 1}$ ,  $\Delta g_\Lambda = [(\Delta p_\Lambda)^T, (\Delta q_\Lambda)^T]^T \in \mathbb{R}^{2|\Lambda| \times 1}$ . The  $\mathbf{R} \in \mathbb{C}^{M \times CD}$  matrix

$$\mathbf{R}_{p_i} = \frac{\partial \Phi(p_\Lambda, q_\Lambda)}{\partial p_i}, \mathbf{R}_{q_i} = \frac{\partial \Phi(p_\Lambda, q_\Lambda)}{\partial q_i}. \quad (36)$$

Assume that  $\Delta p_\Lambda$  and  $\Delta q_\Lambda$  are independent of each other and of the noise. The covariance matrix of  $\Delta g_\Lambda$  is

$$\mathbf{C}_g = \begin{bmatrix} \mathbf{C}_p & 0_{|\Lambda|} \\ 0_{|\Lambda|} & \mathbf{C}_q \end{bmatrix} \in \mathbb{C}^{2|\Lambda| \times 2|\Lambda|}, \quad (37)$$

where  $\mathbf{C}_p = E[\Delta p_\Lambda (\Delta p_\Lambda)^H] \in \mathbb{C}^{|\Lambda| \times |\Lambda|}$ ,  $\mathbf{C}_q = E[\Delta q_\Lambda (\Delta q_\Lambda)^H] \in \mathbb{C}^{|\Lambda| \times |\Lambda|}$ ,  $\mathbf{C}_p^{-1} = \mathbf{D}_p^H \mathbf{D}_p$  and  $\mathbf{C}_q^{-1} = \mathbf{D}_q^H \mathbf{D}_q$ . In the case of RSF radar  $\mathbf{u} = \left[ (\mathbf{D}_p \Delta p_\Lambda)^T, (\mathbf{D}_q \Delta q_\Lambda)^T, \left( \frac{1}{\sigma_w} \mathbf{w} \right)^T \right]^T \in \mathbb{C}^{(2|\Lambda|+M) \times 1}$ , and  $\mathbf{W}_x = [\mathbf{H}_p, \mathbf{H}_q, \sigma_w \mathbf{I}_N] \in \mathbb{C}^M \times (2|\Lambda|+M)$ , where  $\mathbf{H}_p = \mathbf{G}_q (\mathbf{D}_q^{-1} \otimes \mathbf{I}_{|\Lambda|}) (\mathbf{I}_{|\Lambda|} \otimes \mathbf{x}_\Lambda) \in \mathbb{C}^{M \times |\Lambda|}$ ,  $\mathbf{H}_q = \mathbf{G}_q (\mathbf{D}_q^{-1} \otimes \mathbf{I}_{|\Lambda|}) (\mathbf{I}_{|\Lambda|} \otimes \mathbf{x}_\Lambda) \in \mathbb{C}^{M \times |\Lambda|}$ ,  $\mathbf{G}_q = [\mathbf{R}_{q_1}, \dots, \mathbf{R}_{q_{|\Lambda|}}] \in \mathbb{C}^{M \times |\Lambda|^2}$  and  $\mathbf{H}_q = \mathbf{G}_q (\mathbf{D}_q^{-1} \otimes \mathbf{I}_{|\Lambda|}) (\mathbf{I}_{|\Lambda|} \otimes \mathbf{x}_\Lambda) \in \mathbb{C}^{M \times |\Lambda|}$ . Since  $\mathbf{p}$  and  $\mathbf{q}$  are both real, formula (32) is used to update

grid points, in which the imaginary parts of  $\Delta \mathbf{p}$  and  $\Delta \mathbf{q}$  estimates are abandoned.

## 5 Simulations

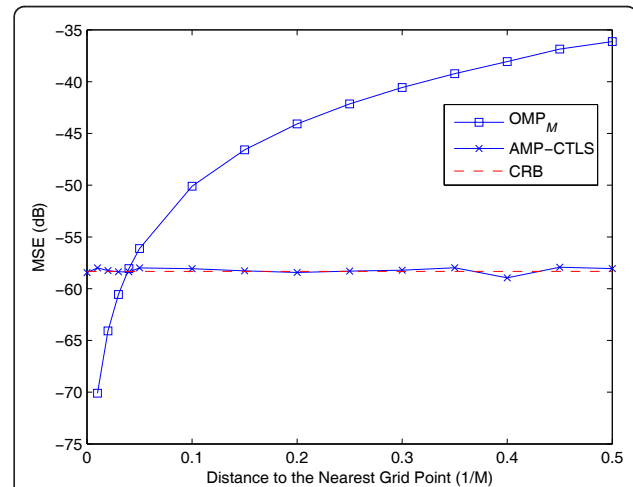
Numerical results are provided to illustrate the performance of the new algorithm. In all examples, the noise is additive Gaussian white noise.

### 5.1 Accuracy of AMP-CTLS

We compare the accuracy of AMP-CTLS with standard OMP. We assume that there is a single sinusoid in the measurements of form (27), where  $\alpha = 1$  and the signal to noise ratio  $\text{SNR} = \alpha^2/\sigma^2 = 5$  dB, where  $\sigma^2$  is the variance of noise. The number of measurements  $M$  is 32. The frequency of sinusoid is varied between two adjoining frequency grid points. The mean square errors (MSEs) of frequency estimates are calculated. The MSEs are compared with the corresponding Cramer-Rao lower bound (CRB) [31]. The frequency is uniformly divided into  $m$  grid points in xMP <sub>$m$</sub>  (OMP <sub>$M$</sub> , OMP<sub>10M</sub>, CoSaMP <sub>$M$</sub> , etc.) and into  $M$  points in AMP-CTLS. AMP-CTLS is configured as follows: IJE loops no more than 14 times; the normalization factors in (11) are  $\mathbf{D} = \mathbf{I}/(\sigma_{\Delta f})$ ,  $\sigma_{\Delta f} = 0.005$ , and  $\sigma_w = 1$ . As shown in Figure 1, MSEs of AMP-CTLS are close to the CRB and lower than those of OMP, except when the sinusoid is in the vicinity of the grid point.

### 5.2 Convergence of AMP-CTLS

We first discuss the convergence speed of the proposed IJE algorithm in noise-free case. Suppose that the sinusoid is located at  $f = 9.5/M$ ,  $M = 32$ . Other conditions

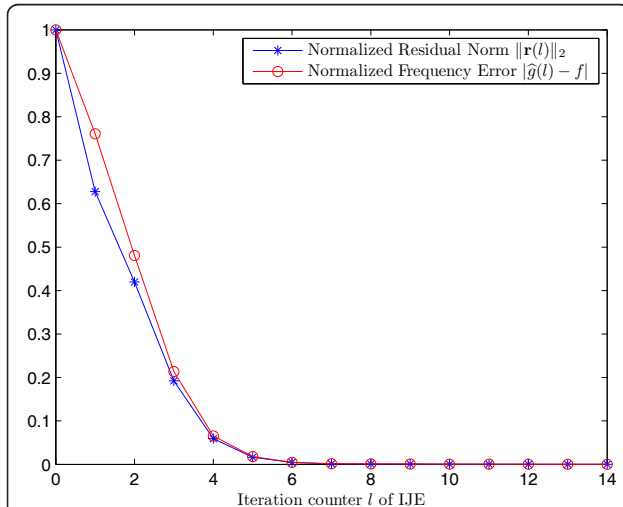


**Figure 1** MSEs of the frequency estimates obtained from 1,000 independent Monte-Carlo trials via OMP and AMP-CTLS. The distance to the nearest grid point is normalized by  $1/M$ . CRB denotes Cramer-Rao bound.

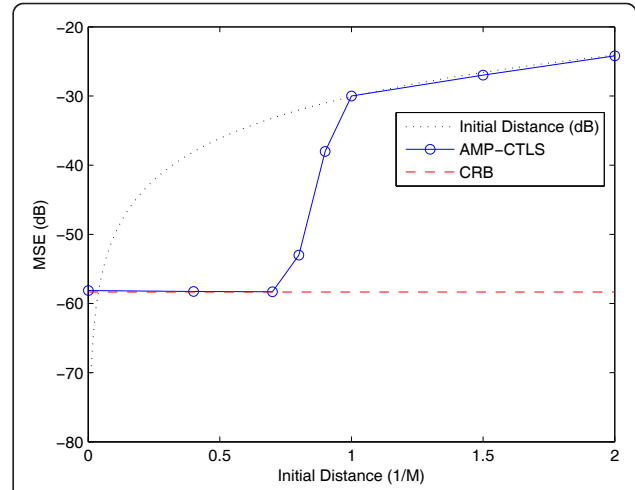
are the same as described in Section 5.1. In the  $l$ th iteration of IJE, we can obtain a grid point  $\hat{g}(l)$  with (5) and residual error  $\mathbf{r}(l) = \mathbf{y} - \Phi(\hat{g})\hat{\mathbf{x}}(l)$  after (6). we calculate the norm of residual  $\|\mathbf{r}(l)\|_2$  and the grid error  $|\hat{g}(l) - f|$ , and normalize the results with  $\|\mathbf{r}(0)\|_2$  and  $|\hat{g}(0) - f|$ , respectively. As shown in Figure 2, both the residual error and the grid error converge fast (about five steps) to 0 in noiseless case.

The purpose of what follows is to discuss the feasible zone of the initial grid points in noisy circumstance. In Section 2.6, the convergence analysis of IJE is based on the assumption that the transform  $\Phi$  is linear. This is only approximately satisfied in harmonic retrieval when the higher order terms of Taylor expansion (8) are ignorable, which means that the grid points indexed in the support set are required to be close to the actual frequencies. We assign SNR = 5 dB and the initial frequency grid point as  $g(0) = 9/M$ . The true frequency of the sinusoid varies from  $9/M$  to  $11/M$ .

As shown in Figure 3, when the distance (normalized by  $1/M$ ) between the true frequency and the initial grid point is less than 0.7, the initial grid is adjusted to be close to the actual value, and MSEs of the frequency estimates converge to CRB. When the distance is greater than 1, the AMP-CTLS curve is close to the initial distance, which means that AMP-CTLS fails to improve the initial grid, because errors of Taylor expansion cannot be ignored and affect convergence of the algorithm.



**Figure 2** The normalized errors at different IJE iterations in noiseless case.  $\|\mathbf{r}(l)\|_2$  denotes the norm of residual error after  $l$ th iteration.  $f$  is actual frequency of the sinusoid and  $\hat{g}(l)$  is the estimate of grid point. In the plots,  $\|\mathbf{r}(l)\|_2$  and  $|\hat{g}(l) - f|$  are normalize with  $\|\mathbf{r}(0)\|_2$  and  $|\hat{g}(0) - f|$ , respectively.



**Figure 3** MSEs of the frequency estimates versus distances between the initial grid point and the actual values. The distance is normalized by  $1/M$ . The estimates are obtained from 1,000 independent Monte-Carlo trials. 'Initial distance (dB)' in the legend denotes the square differences between the frequency grid point and the true ones.

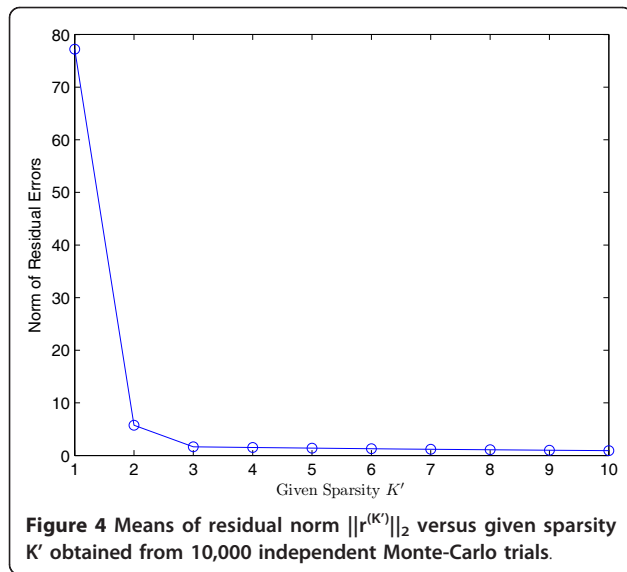
### 5.3 Input of sparsity

In Sections 5.1 and 5.2, we assume that the sparsity  $K$ , i.e., the number of modes, is known and we use  $K$  to terminate AMP-CTLS, while a priori sparsity is not obligatory. When  $K$  is unknown, we can use norm of residual error  $\mathbf{r} = \mathbf{y} - \Phi(\mathbf{g}_\Lambda) \mathbf{x}_\Lambda$  as termination criterion.

Furthermore, AMP-CTLS does not seriously rely on the given sparsity  $K'$ , and the performance is slightly affected when  $K' > K$ . Suppose there are three sinusoids denoted as  $\text{Si}_1$ ,  $\text{Si}_2$ , and  $\text{Si}_3$ , where  $\alpha_1 = 20$ ,  $\alpha_2 = 15$ ,  $\alpha_3 = 1$ ,  $f_1 = 3.15/M$ ,  $f_2 = 4.2/M$ ,  $f_3 = 7.25/M$ ,  $M = 32$ .  $\text{SNR}_3 = \alpha_3^2/\sigma^2 = 10$  dB. In AMP-CTLS, frequency is uniformly grided to  $2M$  points, and other configurations are the same as described in Section 5.1.

We calculate means of the final residual norm  $\|\mathbf{r}^{(K')}\|_2$  versus  $K'$  and present the results in Figure 4. When all of the sinusoids have been chosen into the support set and  $K' \geq K$ , energy of the sinusoids are canceled thoroughly and only noise exist in the residual. The norm of residual error becomes small and is slowly reduced along with  $K'$ . The results illustrate that we can use threshold of values or decrease rate of the norm of residual to end AMP-CTLS loops.

Spurious sinusoids emerge when  $K' > K$ . Denote the amplitude estimates by  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{K'}$  in descend order of magnitudes and their counterparts of frequency estimates by  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_{K'}$ . MSEs of  $f_3$  estimates versus  $K'$  are presented in Figure 5, which indicates that accuracy of frequency estimates of  $\text{Si}_3$  is slightly affected ( $< 2$  dB) by  $K'$ .

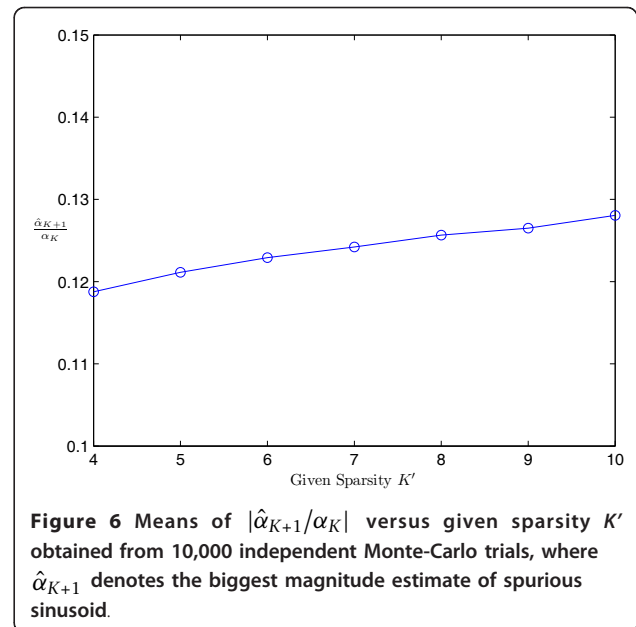


We also calculate  $|\hat{\alpha}_{K+1}/\alpha_K|$  as measurement of the level of spurious sinusoids. Figure 6 presents the results and shows that the ratios  $|\hat{\alpha}_{K+1}/\alpha_K|$  stay at low level ( $<0.2$ ) and are not sensitive to  $K'$ .

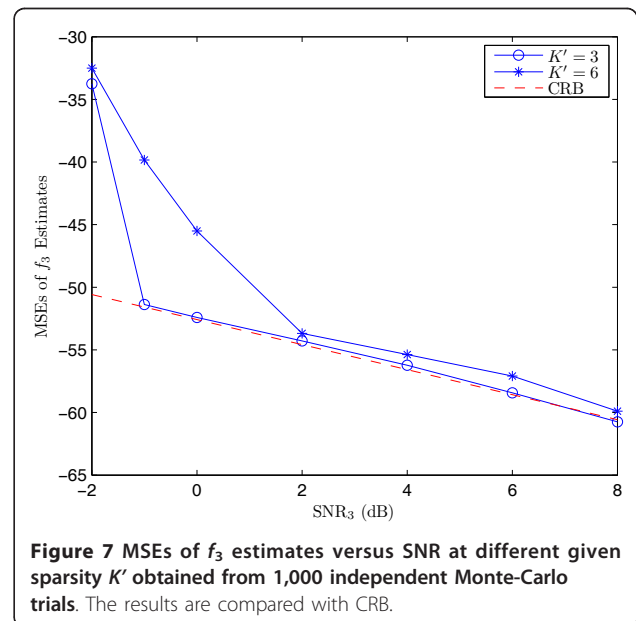
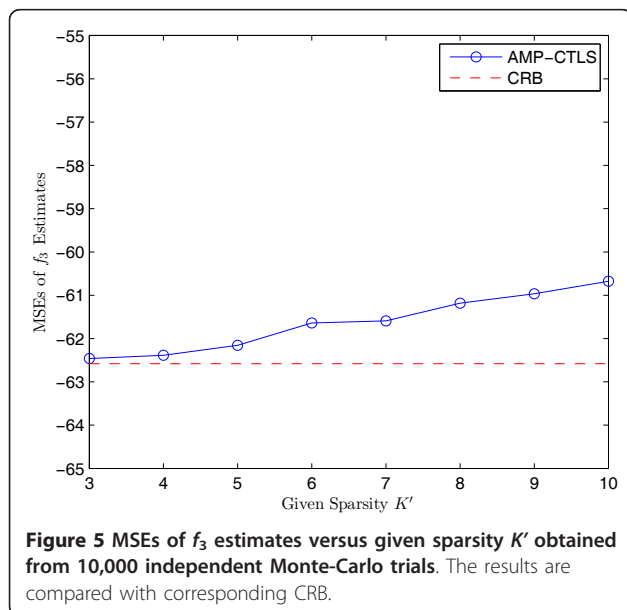
Figure 7 presents MSEs of  $f_3$  estimates versus SNR at different  $K'$ . Noise variance  $\sigma^2$  is altered such that  $\text{SNR}_3$  varies. The MSEs converges to CRB at high SNR ( $\text{SNR}_3 > 2$  dB) when  $K' = K = 3$ . The results of  $K' = 6$  are close to those of  $K' = 3$ .

#### 5.4 Recovering small sinusoids

We compare the performance on recovering weak sinusoids of AMP-CTLS with CS methods, e.g., OMP and



CoSaMP, and conventional spectral analysis methods, e.g., ESPRIT and root MUSIC [18]. Suppose there are three sinusoids denoted as  $S_{i1}$ ,  $S_{i2}$ , and  $S_{i3}$ , where  $\alpha_1 = 20$ ,  $\alpha_2 = 15$ ,  $\alpha_3 = 1$ ,  $f_1 = 3.15/M$ ,  $f_2 = 5.2/M$ ,  $f_3 = 3.95/M$ ,  $M = 32$ .  $\text{SNR}_3 = \alpha_3^2/\sigma^2 = 5$  dB. The number of sinusoids  $K = 3$  is assumed to be known. CoSaMP iterates 50 times. In both ESPRIT and root MUSIC, the model orders are set as  $K$ , and the covariance matrix orders are  $M/2$  according to [32]. ESPRIT and root MUSIC output frequency estimates, and the corresponding magnitudes are obtained by projection on these frequencies.





AMP-CTLS is configured the same as mentioned in Section 5.3.

Some intuitive results are presented in Figure 8. The  $Si_3$  is recovered by the AMP-CTLS algorithm and is masked via other tested algorithms. CoSaMP<sub>100M</sub> is also tested, but the results are not displayed because the amplitude estimates are too large ( $>1,000$ ), which is caused by projection onto the ill-conditioned matrix consisting of highly correlated atoms. In OMP<sub>2M</sub>, the sinusoids are not exactly at the grid points, so the energies of  $Si_1$  and  $Si_2$  cannot be totally canceled in the beginning two iterations, and the leakage of the energies masks the smallest signal  $Si_3$ . In OMP<sub>100M</sub>, all sinusoids are placed at the grid points, and  $Si_1$  and  $Si_2$  are better recovered than in OMP<sub>2M</sub>, but energy leakage of dominant sinusoids still exists. In AMP-CTLS, the grid points are adaptively adjusted to match the sinusoids, so the algorithm is less sensitive to grid mismatch and can achieve better performance than OMP and CoSaMP even if the frequency space is sparsely divided. ESPRIT and root MUSIC do not correctly recover  $Si_3$  as AMP-CTLS does. Since there is only one snapshot data, smoothing method [30] is used in these two methods to estimate the covariance matrix, which results in aperture loss [18].

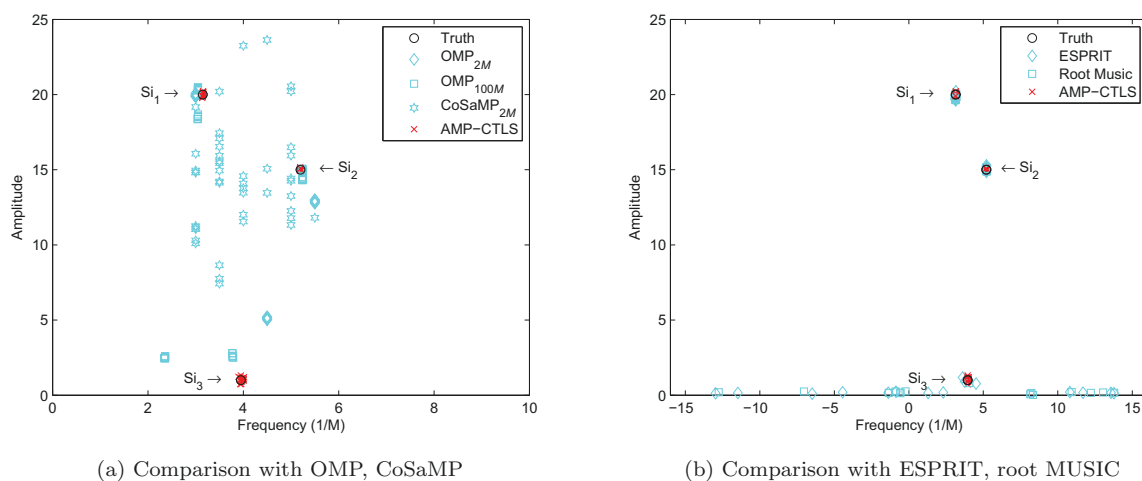
### 5.5 Range-velocity joint estimate in RSF radar

In this section, we discuss merits of AMP-CTLS in recovering small targets with RSF radar. Suppose there are three targets: two large targets  $T_1$  and  $T_2$  and a small target  $T_3$ . The number of measurements  $M$  is 32.

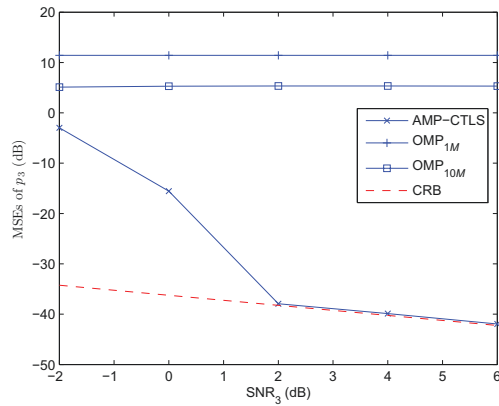
The scattering intensities are  $\alpha_1 = \alpha_2 = 10$ ,  $\alpha_3 = 1$ , and the ranges and the velocities are set such that the  $p, q$  parameters are  $p_1 = 10.1/M$ ,  $p_2 = 10.7/M$ ,  $p_3 = 20/M$ ,  $q_1 = 19.4/M$ ,  $q_2 = 10.2/M$  and  $q_3 = 15.2/M$ . AMP-CTLS is configured as follows: the  $p, q$  spaces are both uniformly divided into  $M$  grid points; the normalization factors are  $\mathbf{D}_p = \mathbf{D}_q = \mathbf{I}/\sigma_\Delta$ ,  $\sigma_\Delta = 0.025$ ,  $\sigma_w = 1$ ; and the IJE algorithm iterates fewer than 14 times. In OMP<sub>m</sub>, both the  $p$  and  $q$  spaces are uniformly divided into  $m$  grid points. Note that all of the targets lie on the grid points in OMP<sub>100M</sub>. We focus on the results of recovering the weakest target  $T_3$ . Change the noise covariance  $\sigma^2$ ; thus, the signal to noise ratio  $\text{SNR}_3 = \alpha_3^2/\sigma^2$  varies. Calculate MSEs of  $p_3, q_3$  parameters. As shown in Figure 9, the MSEs with AMP-CTLS are lower than those with OMP and converge to the CRB when the  $\text{SNR}_3$  is no less than 2 dB. The difference between these MSEs of AMP-CTLS at high SNR and CRB is less than 0.5 dB.

### 5.6 DoA estimation

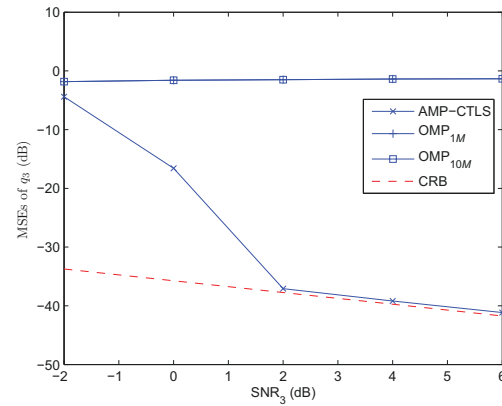
In this section, AMP-CTLS is compared with the Lasso-based TLS method WSS-TLS [14] on direction of arrival (DoA) estimation. The goal is estimating DoA of plane waves from far-field, narrowband sources with uniform linear array of antennas [14]. We focus on the single-snapshot case. Suppose the antenna array contains  $M = 8$  elements and the interval between neighboring elements  $d = 1/2$  wavelength. There are two sources ( $K = 2$ ) from angles  $\theta_1 = -29^\circ$  and  $\theta_2 = 13^\circ$ . The amplitudes



**Figure 8** Amplitude and frequency estimates of the sinusoids obtained with OMP, CoSaMP, ESPRIT, root MUSIC and AMP-CTLS from 15 independent Monte-Carlo trials. The estimates are compared with the true parameters. Subscripts of OMP and CoSaMP denote numbers of grid points. Note that in OMP<sub>100M</sub>, all of the sinusoids exactly lie on the grid points. The results of conventional MP methods and spectral analysis methods are, respectively, shown in (a) and (b) for better readability.



(a) MSEs of  $p_3$  estimates



(b) MSEs of  $q_3$  estimates

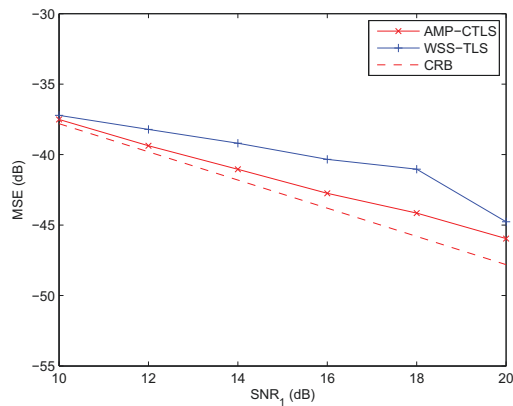
**Figure 9** MSEs of  $p_3$  and  $q_3$  estimates of OMP and AMP-CTLS versus  $\text{SNR}_3$  obtained from 1,000 independent Monto-Carlo trials. The results are compared with CRB.

$\alpha_1 = \alpha_2 = 1$  and  $\text{SNR}_1 = \text{SNR}_2 = \alpha_1^2/\sigma^2$ , where  $\sigma^2$  is the noise variance. The angle space from  $-90^\circ$  to  $90^\circ$  are uniformly divided to  $N = 90$  grid points; thus both sources are  $1^\circ$  off the nearest grid points. The WSS-TLS algorithm is set according to [14]. Since WSS-TLS returns multiple nonzero DoA estimates, we choose two peaks with largest magnitudes as the estimates of  $\theta_1$  and  $\theta_2$ . AMP-CTLS models DoA estimation as an harmonic retrieval problem and outputs frequency estimates  $\hat{f} \in [0, 1)$ . Denote  $\tilde{f} = \hat{f} - 0.5(\text{sgn}(\hat{f} - 0.5) + 1)$ , where  $\text{sgn}(\cdot)$  represents signum function; thus the DoA estimate with AMP-CTLS is obtained by  $\hat{\theta} = \arcsin^{-1}(\tilde{f}/d)$ . In AMP-CTLS, IJE loops no more than 50 times; the normalization factors in (11) are  $\mathbf{D} = \mathbf{I}/(\sigma_{\Delta f})$ ,  $\sigma_{\Delta f} = 0.005$ ,  $\sigma_w = 1$ . MSEs of  $\theta_1$  and  $\theta_2$  estimates versus SNR are shown in Figure 10a, b,

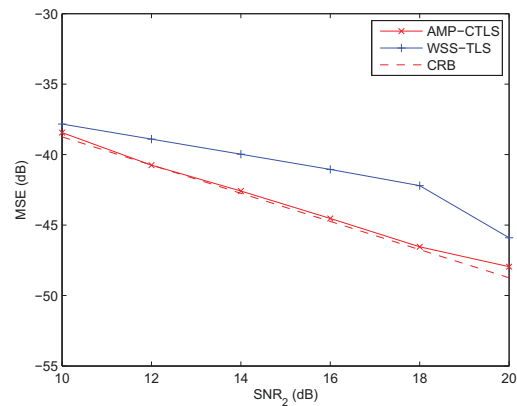
respectively. The results indicate that MSEs of AMP-CTLS are closer to CRB than those of WSS-TLS.

## 6 Conclusion

To alleviate the off-grid problem in grid-based MP methods, we implement CTLS into the OMP framework and propose a new algorithm, namely AMP-CTLS. Unlike traditional MP methods, AMP-CTLS adaptively adjusts the grid and dictionary. The convergence of AMP-CTLS is analyzed, and the initial conditions of the algorithm are discussed. Numerical examples indicate that the advantages of AMP-CTLS over OMP and CoSaMP are twofold: (1) it is still efficient even when the continuous parameter space is sparsely divided, but OMP or CoSaMP suffers from performance degradation when the space is not divided reasonably; (2) it can achieve higher accuracy, and the MSEs converge to CRB.



(a) MSEs of  $\theta_1 = -29^\circ$  estimates



(b) MSEs of  $\theta_2 = 13^\circ$  estimates

**Figure 10** MSEs of  $\theta_1$  and  $\theta_2$  estimates of WSS-TLS and AMP-CTLS obtained from 1,000 independent Monto-Carlo trials.

## Appendix 1

We recall the complex version of Newton's method in [20]. Minor changes are made because of differences in notation. The recursion formula of the complex Newton method for (17) is:

$$\mathbf{x}_{\text{CTLS}}^{(t+1)} = \mathbf{x}_{\text{CTLS}}^{(t)} + (\mathbf{A}^* \mathbf{B}^{-1} \mathbf{A} - \mathbf{B}^*)^{-1} (\mathbf{a}^* - \mathbf{A}^* \mathbf{B}^{-1} \mathbf{a}), \quad (38)$$

where

$$\mathbf{h} = (\mathbf{W}_x \mathbf{W}_x^H)^{-1} (\Phi_{\Lambda} \mathbf{x}_{\text{CTLS}}^{(t)} - \mathbf{y}), \quad (39)$$

$$\hat{\mathbf{u}} = -\mathbf{W}_x^H \mathbf{h}, \quad (40)$$

$$\tilde{\mathbf{B}} = \Phi_{\Lambda} + \mathbf{G}(\mathbf{D}^{-1} \otimes \mathbf{I}_{|\Lambda|})(\hat{\mathbf{u}} \otimes \mathbf{I}_{|\Lambda|}), \quad (41)$$

$$\mathbf{Q}_j = \left[ [(\mathbf{R}_1)_{(j)}, \dots, (\mathbf{R}_{|\Lambda|})_{(j)}] \mathbf{D}^{-1}, \mathbf{0}_N \right], j \leq |\Lambda|, \quad (42)$$

$$\tilde{\mathbf{Q}} = [\mathbf{Q}_1^H \mathbf{h}, \mathbf{Q}_2^H \mathbf{h}, \dots, \mathbf{Q}_{|\Lambda|}^H \mathbf{h}], \quad (43)$$

$$\mathbf{a} = (\mathbf{h}^H \tilde{\mathbf{B}})^T, \quad (44)$$

$$\mathbf{A} = -\tilde{\mathbf{Q}}^H \mathbf{W}_x^{\dagger} \tilde{\mathbf{B}} - (\tilde{\mathbf{Q}}^H \mathbf{W}_x^{\dagger} \tilde{\mathbf{B}})^T, \quad (45)$$

$$\mathbf{B} = \tilde{\mathbf{Q}}^H (\mathbf{W}_x^{\dagger} \mathbf{W}_x - \mathbf{I}) \tilde{\mathbf{Q}} + (\tilde{\mathbf{B}}^H (\mathbf{W}_x \mathbf{W}_x^H)^{-1} \tilde{\mathbf{B}})^T \quad (46)$$

## Appendix 2

When the frequencies are constrained to be real, the CTLS solver becomes more complex. Some notations are introduced for simplicity as follows:  $\mathbf{W}_1 = \mathbf{H}$ ,  $\mathbf{W}_2 = \sigma_w \mathbf{I}_N$ ,  $\mathbf{u} = [\mathbf{u}_1^T, \mathbf{u}_2^T]^T$ ,  $\mathbf{z} = \mathbf{y} - \Phi_{\Lambda} \mathbf{x}_{\Lambda}$ . Notice that the matrix  $\mathbf{H}$  is relative to  $\mathbf{x}_{\Lambda}$ . Replace the optimum problem in (12), (14) with:

$$\hat{\mathbf{u}}, \mathbf{x}_{\text{CTLS}} = \arg \min_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{x}_{\Lambda}} \mathbf{u}_1^T \mathbf{u}_1 + \mathbf{u}_2^H \mathbf{u}_2, \quad (47)$$

$$\text{s.t.} \begin{cases} \mathbf{z} - [\mathbf{W}_1, \mathbf{W}_2] \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix} = \mathbf{0}. \\ \mathbf{u}_1^* = \mathbf{u}_1 \end{cases} \quad (48)$$

First, suppose  $\mathbf{x}_{\Lambda}$  is known, and seek the solution of  $\mathbf{u}$ . If both  $\mathbf{W}_1$  and  $\mathbf{W}_2$  are of full-row rank, we have

$$\mathbf{u}_1 = 2\text{Re}(\mathbf{W}_1^H \mathbf{v}), \quad (49)$$

$$\mathbf{u}_2 = 2\mathbf{W}_2^H \mathbf{v}, \quad (50)$$

where

$$\mathbf{v} = -(\mathbf{C}_2^{-1} \mathbf{C}_1 - (\mathbf{C}_1^{-1} \mathbf{C}_2)^*)^{-1} (\mathbf{C}_2^{-1} \mathbf{z} - (\mathbf{C}_1^{-1} \mathbf{z})^*), \quad (51)$$

$$\mathbf{C}_1 = \mathbf{W}_1 \mathbf{W}_1^H + 2\mathbf{W}_2 \mathbf{W}_2^H, \quad (52)$$

$$\mathbf{C}_2 = \mathbf{W}_1 \mathbf{W}_1^T. \quad (53)$$

The solution  $\mathbf{u}_1, \mathbf{u}_2$  depends on  $\mathbf{x}_{\Lambda}$ . Then, calculate  $\mathbf{x}_{\Lambda}$  as

$$\mathbf{x}_{\text{CTLS}} = \arg \min_{\mathbf{x}_{\Lambda}} \mathbf{u}_1^T \mathbf{u}_1 + \mathbf{u}_2^H \mathbf{u}_2. \quad (54)$$

However, it is difficult to solve (54), because the Jacobi matrix and Hessian matrix of  $\mathbf{u}_1^T \mathbf{u}_1 + \mathbf{u}_2^H \mathbf{u}_2$  versus  $\mathbf{x}_{\Lambda}$  are complex. In this article, we simply consider the frequencies to be complex and ignore the imaginary parts.

**Proof:** We prove that when  $\mathbf{x}_{\Lambda}$  is known, the solution of  $\mathbf{u}$  is given as (49) to (53). The Lagrangian

$$L(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \mathbf{u}_1^T \mathbf{u}_1 - \mathbf{v}^H (\mathbf{z} + \mathbf{W}_1 \mathbf{u}_1 + \mathbf{W}_2 \mathbf{u}_2) + \frac{1}{2} \mathbf{u}_2^H \mathbf{u}_2 - \mathbf{v}^T (\mathbf{z}^* + \mathbf{W}_1^* \mathbf{u}_1 + \mathbf{W}_2^* \mathbf{u}_2^*) \quad (55)$$

can be expressed as follows:

$$L(\mathbf{u}, \mathbf{v}) = \frac{1}{2} (\mathbf{u}_1 - 2\text{Re}(\mathbf{W}_1^H \mathbf{v}))^T (\mathbf{u}_1 - 2\text{Re}(\mathbf{W}_1^H \mathbf{v})) - 2(\text{Re}(\mathbf{W}_1^H \mathbf{v}))^T \text{Re}(\mathbf{W}_1^H \mathbf{v}) - 2\mathbf{v}^H \mathbf{W}_2 \mathbf{W}_2^H \mathbf{v} + \frac{1}{2} (\mathbf{u}_2 - 2\mathbf{W}_2^H \mathbf{v})^H (\mathbf{u}_2 - 2\mathbf{W}_2^H \mathbf{v}) - 2\text{Re}[\mathbf{v}^H \mathbf{z}]. \quad (56)$$

When  $\mathbf{u}_1 - 2\text{Re}(\mathbf{W}_1^H \mathbf{v})$  and  $\mathbf{u}_2 = 2\mathbf{W}_2^H \mathbf{v}$ , the Lagrangian reaches the infimum; thus, the Lagrange dual function is obtained as

$$\gamma(\mathbf{v}) = \inf_{\mathbf{u}} L(\mathbf{u}, \mathbf{v}) = -4\mathbf{v}^T \mathbf{W}_2^* \mathbf{W}_2^T \mathbf{v}^* - 4\text{Re}(\mathbf{v}^H \mathbf{z}) - (\mathbf{v}^H \mathbf{W}_1 + \mathbf{v}^T \mathbf{W}_1^*) (\mathbf{W}_1^H \mathbf{v} + \mathbf{W}_1^T \mathbf{v}^*). \quad (57)$$

Calculate the Jacobi and Hessian matrix of  $\gamma(\mathbf{v})$ :

$$\frac{\partial \gamma}{\partial \mathbf{v}^*} = -\frac{1}{2} \mathbf{W}_1 \mathbf{W}_1^H \mathbf{v} - \frac{1}{2} \mathbf{W}_1 \mathbf{W}_1^T \mathbf{v}^* - \mathbf{W}_2 \mathbf{W}_2^H \mathbf{v} - \frac{1}{2} \mathbf{z}, \quad (58)$$

$$\frac{\partial^2 \gamma}{\partial \mathbf{v}^T \partial \mathbf{v}^*} = -\frac{1}{2} \mathbf{W}_1 \mathbf{W}_1^H - \mathbf{W}_2 \mathbf{W}_2^H. \quad (59)$$

Because  $\mathbf{W}_1, \mathbf{W}_2$  are of full-row rank,  $\frac{\partial^2 \gamma}{\partial \mathbf{v}^T \partial \mathbf{v}^*}$  is a negative definite matrix. We solve  $\mathbf{v}$  with the optimum condition  $\frac{\partial \gamma}{\partial \mathbf{v}^*} = \mathbf{0}$ , and obtain (51) to (53). The proof is complete.

## Algorithm 1. The CTLS technique

- 1) Input the dictionary  $\Phi_{\Lambda}$  and all the coefficient matrices  $\mathbf{R}_i$ .
- 2) Compose  $\mathbf{W}_x$  and solve (17) with the initial value given in (20).
- 3) Calculate  $\hat{\mathbf{u}}$  via (18) and (19).

4) Extract  $\widehat{\Delta \mathbf{g}}_{\Lambda}$  from  $\hat{\mathbf{u}}$ .

### Algorithm 2. The AMP-CTLS algorithm

1) Divide the continuous parameter  $f$  into grid point  $\hat{\mathbf{g}}^{(0)}$ ; input the sparsity level  $K$  or residual threshold  $\delta$ .

Set the support set  $\Lambda^{(0)} = \emptyset$ , and the residual error  $\mathbf{r}^{(0)} = \mathbf{y}$ .

2) Calculate the correlations  $p_i = \left\langle \mathbf{r}^{(k)}, \Phi \left( \hat{\mathbf{g}}_i^{(k)} \right) \right\rangle$ .

3) Find the index  $n = \arg \max |p_i|$ .

4) Merge the support set  $\Lambda^{(k+1)} = \Lambda^{(k)} \cup \{n\}$ .

5) Solve (3) with the IJE algorithm. Then we get  $\hat{\mathbf{g}}_{\Lambda}^{(k+1)}$  and  $\hat{\mathbf{x}}_{\Lambda}^{(k+1)}$ .

6) Update the residual error  $\mathbf{r}^{(k+1)} = \mathbf{y} - \Phi \left( \hat{\mathbf{g}}_{\Lambda}^{(k+1)} \right) \hat{\mathbf{x}}_{\Lambda}^{(k+1)}$ .

7) Increase  $k$ . Return to Step 2 until stop criterion, e.g.,  $k < K$ ,  $\|\mathbf{r}^{(k)}\|_2 < \delta$  or  $\|\mathbf{r}^{(k)}\|_2 < \delta \|\mathbf{r}^{(k-1)}\|_2$ , is satisfied.

8) Simultaneously output  $\hat{\mathbf{x}}_{\Lambda}^{(k)}$  and  $\hat{\mathbf{g}}_{\Lambda}^{(k)}$ , and set the elements of  $\mathbf{x}$  not indexed in  $\Lambda$  to 0.

### Acknowledgements

This study was supported in part by the National Natural Science Foundation of China (No. 40901157) and in part by the National Basic Research Program of China (973 Program, No. 2010CB731901). Thanks to the anonymous reviewers for many valuable comments and to Hao Zhu for helpful discussions and her Matlab® programs of WSS-TLS.

### Competing interests

The authors declare that they have no competing interests.

Received: 12 September 2011 Accepted: 4 April 2012

Published: 4 April 2012

### References

- RG Baraniuk, Compressive sensing [lecture notes]. *IEEE Signal Process Mag.* **24**(4), 118–121 (2007)
- W Dai, O Milenkovic, Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans Inf Theory.* **55**(5), 2230–2249 (2009)
- SG Mallat, Z Zhang, Matching pursuits with time-frequency dictionaries. *IEEE Trans Signal Process.* **41**(12), 3397–3415 (1993). doi:10.1109/78.258082
- MA Davenport, MB Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans Inf Theory.* **56**(9), 4395–4401 (2010)
- D Needell, R Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found Comput Math.* **9**(3), 317–334 (2009). doi:10.1007/s10208-008-9031-3
- D Donoho, I Drori, Y Tsaig, J Starck, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit, (Department of Statistics, Stanford University, California, 2006)
- D Needell, J Tropp, CoSaMP, Iterative signal recovery from incomplete and inaccurate samples. *Appl Comput Harmonic Anal.* **26**(3), 301–321 (2009). doi:10.1016/j.acha.2008.07.002
- T Huang, Y Liu, H Meng, X Wang, Randomized step frequency radar with adaptive compressed sensing. In: *Proc IEEE Radar Conf (RADAR)*, Kansas City, Missouri, USA 411–414 (2011)
- S Shah, Y Yu, A Petropulu, Step-frequency radar with compressive sampling (SFR-CS). In: *Proc IEEE Int Acoustics Speech and Signal Processing (ICASSP) Conf*, Dallas, Texas, USA 1686–1689 (2010)
- MM Hyder, Direction-of-arrival estimation using a mixed  $\ell_{2,0}$  norm approximation. *IEEE Trans Signal Process.* **58**(9), 4646–4655 (2010)
- C Zheng, G Li, H Zhang, X Wang, An approach of regularization parameter estimation for sparse signal recovery. In: *Proc IEEE 10th Int Signal Processing (ICSP) Conf*, Beijing, China 385–388 (2010)
- C Zheng, G Li, H Zhang, X Wang, An approach of DOA estimation using noise subspace weighted  $\ell_1$  minimization. In: *Proc IEEE Int Acoustics, Speech and Signal Processing (ICASSP) Conf*, Prague, Czech 2856–2859 (2011)
- Y Chi, LL Scharf, A Pezeshki, AR Calderbank, Sensitivity to basis mismatch in compressed sensing. *IEEE Trans Signal Process.* **59**(5), 2182–2195 (2011)
- H Zhu, G Leus, GB Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling. *IEEE Trans Signal Process.* **59**(5), 2002–2016 (2011)
- DH Chae, P Sadeghi, RA Kennedy, Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals. In: *Proc 2nd Int Future Computer and Communication (ICFCC) Conf*, Wuhan, China. **2**, V2.739–V2.743 (2010)
- SD Cabrera, S Malladi, R Mulpuri, AE Brito, Adaptive refinement in maximally sparse harmonic signal retrieval. In: *IEEE 11th Proc and the 3rd IEEE Signal Processing Education Workshop Digital Signal Processing Workshop*, Taos Ski Valley, New Mexico, USA 231–235 (2004)
- G Peyre, Best basis compressed sensing. *IEEE Trans Signal Process.* **58**(5), 2613–2622 (2010)
- P Stoica, R Moses, *Spectral Analysis of Signals*, Pearson/Prentice Hall, Upper Saddle River, (2005)
- R Bellman, *Introduction to Matrix Analysis*, (Society for Industrial Mathematics, Philadelphia, 1997)
- TJ Abatzoglou, JM Mendel, GA Harada, The constrained total least squares technique and its applications to harmonic superresolution. *IEEE Trans Signal Process.* **39**(5), 1070–1087 (1991). doi:10.1109/78.80955
- G Golub, C Van Loan, An analysis of the total least squares problem. *SIAM J Numer Anal.* **17**(6), 883–893 (1980). doi:10.1137/0717073
- P Boufounos, MF Duarte, RG Baraniuk, Sparse signal reconstruction from noisy compressive measurements using cross validation. In: *Proc IEEE/SP 14th Workshop Statistical Signal Processing SSP'07*, Madison, Wisconsin, USA 299–303 (2007)
- AE Brito, SD Cabrera, C Villalobos, Optimal sparse representation algorithms for harmonic retrieval. In: *Proc Conf Signals, Systems and Computers Record of the Thirty-Fifth Asilomar Conf*, Pacific Grove, California, USA. **2**, 1407–1411 (2001)
- BD Rao, K Kreutz-Delgado, An affine scaling methodology for best basis selection. *IEEE Trans Signal Process.* **47**, 187–200 (1999). doi:10.1109/78.738251
- IF Gorodnitsky, BD Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm. *IEEE Trans Signal Process.* **45**(3), 600–616 (1997). doi:10.1109/78.558475
- S Cabrera, J Rosiles, A Brito, Affine scaling transformation algorithms for harmonic retrieval in a compressive sampling framework. In: *Proc Wavelets XII, SPIE*, San Diego, California, USA, **6701**, 67012D.1–67012D.12 (2007)
- JA Tropp, Greed is good: algorithmic results for sparse approximation. *IEEE Trans Inf Theory.* **50**(10), 2231–2242 (2004). doi:10.1109/TIT.2004.834793
- SRJ Axelsson, Analysis of random step frequency radar and comparison with experiments. *IEEE Trans Geosci Remote Sens.* **45**(4), 890–904 (2007)
- Y Liu, H Meng, G Li, X Wang, Range-velocity estimation of multiple targets in randomised stepped-frequency radar. *Electron Lett.* **44**(17), 1032–1034 (2008). doi:10.1049/el:20081608
- JW Odendaal, E Barnard, CWI Pistorius, Two-dimensional superresolution radar imaging using the MUSIC algorithm. *IEEE Trans Antennas Propag.* **42**(10), 1386–1391 (1994). doi:10.1109/8.320744
- SF Yau, Y Bresler, A compact Cramer-Rao bound expression for parametric estimation of superimposed signals. *IEEE Trans Signal Process.* **40**(5), 1226–1230 (1992). doi:10.1109/78.134484
- K Mahata, T Soderstrom, ESPRIT-like estimation of real-valued sinusoidal frequencies. *IEEE Trans Signal Process.* **52**(5), 1161–1170 (2004). doi:10.1109/TSP.2004.826169

doi:10.1186/1687-6180-2012-76

Cite this article as: Huang et al.: Adaptive matching pursuit with constrained total least squares. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:76.